CSE- 1006 LAB Assignment

Academic year: 2020-2021 **Semester:** FALL

Faculty Name: Dr. Arun kumar Gopu **Date:** 16/3/2022

Student name: M.Taran **Reg. no.:** 19BCE7346

Finding Duplicate Values:

duplicated() determines which elements of a vector or data frame are duplicates of elements with smaller subscripts, and returns a logical vector indicating which elements (rows) are duplicates.

anyDuplicated() returns the index of the first duplicate value if any, otherwise 0. anyDuplicated() is a "generalised" more efficient shortcut for any(duplicated())

```
> anyDuplicated(x)
[1] 18
>
```

EXERCISES

· Create a vector as x <- c(9:20, 1:5, 3:7, 0:8)

```
> x <- c(9:20, 1:5, 3:7, 0:8)
> x
 [1]  9 10 11 12 13 14 15 16 17 18 19 20
[13]  1  2  3  4  5  3  4  5  6  7  0  1
[25]  2  3  4  5  6  7  8
```

· Use duplicated() function to print the logical vector indicating the duplicate values present in x

```
> duplicated(x)
 [1] FALSE FALSE FALSE FALSE FALSE FALSE
 [7] FALSE FALSE FALSE FALSE FALSE FALSE
[13] FALSE FALSE FALSE FALSE FALSE  TRUE
[19]  TRUE  TRUE FALSE FALSE FALSE  TRUE
[25]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[31] FALSE
```

· Observe the output of duplicated(x, fromLast = TRUE)

```
> duplicated(x, fromLast = TRUE)
 [1] FALSE FALSE FALSE FALSE FALSE FALSE
 [7] FALSE FALSE FALSE FALSE FALSE FALSE
[13]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[19]  TRUE  TRUE  TRUE  TRUE FALSE FALSE
[25] FALSE FALSE FALSE FALSE FALSE FALSE
[31] FALSE
```

· What is the difference between duplicated(x) and duplicated(x,fromLast=TRUE)

· Extract duplicate elements from x

```
> x[duplicated(x)]
 [1] 3 4 5 1 2 3 4 5 6 7
>
```

· Extract unique elements from x

```
>
> x[!duplicated(x)]
 [1]  9 10 11 12 13 14 15 16 17 18 19 20  1  2  3  4  5
[18]  6  7  0  8
>
```

· Print duplicate elements from x in different order (Hint: Use duplicated(x, fromLast = TRUE))

```
>
> x[duplicated(x,fromLast = TRUE)]
 [1] 1 2 3 4 5 3 4 5 6 7
>
```

· Extract unique elements from x in different order (Hint: Use duplicated(x, fromLast = TRUE))

```
>
> x[!duplicated(x,fromLast = TRUE)]
 [1]  9 10 11 12 13 14 15 16 17 18 19 20  0  1  2  3  4
[18]  5  6  7  8
>
>
```

· Print the indices of duplicate elements

```
>
> which(duplicated(x))
 [1] 18 19 20 24 25 26 27 28 29 30
>
>
```

· Print the indices of unique elements

```
>
> which(!duplicated(x))
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17
[18] 21 22 23 31
>
```

· How many unique elements are in x

```
>
> sum(!duplicated(x))
[1] 21
>
```

· How many duplicate elements are in x

```
> sum(duplicated(x))
[1] 10
>
```

· Create a dataframe df :

a <- c(rep("A", 3), rep("B", 3), rep("C",2))

b <- c(1,1,2,4,1,1,2,2)

df <-data.frame(a,b)

```
>
> a <- c(rep("A", 3), rep("B", 3), rep("C",2))
> b <- c(1,1,2,4,1,1,2,2)
> df <-data.frame(a,b)
>
>
```

· Use duplicated() function to print the logical vector indicating the duplicate values present in dataframe "df"

```
>
> duplicated(df)
[1] FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE
[8]  TRUE
>
```

· Extract duplicate elements from dataframe "df"

```
>
> df[duplicated(df),]
  a b
2 A 1
6 B 1
8 C 2
>
>
```

· Extract unique elements from dataframe "df"

```
>
> df[!duplicated(df),]
  a b
1 A 1
3 A 2
4 B 4
5 B 1
7 C 2
>
```

· Print the indices of duplicate elements

```
>
> which(duplicated(df))
[1] 2 6 8
>
```

· Print the indices of unique elements

```
>
> which(!duplicated(df))
[1] 1 3 4 5 7
>
```

· How many unique elements are in dataframe "df"

```
>
> sum(!duplicated(df))
[1] 5
>
>
```

· How many duplicate elements are in dataframe "df

```
>
> sum(duplicated(df))
[1] 3
>
```